Methylation Markers for Small Cell Lung Cancer in Peripheral Blood Leukocyte DNA

Liang Wang, MD, Jeremiah A. Aakre, BS,† Ruoxiang Jiang, BS,† Randolph S. Marks, MD,‡ Yanhong Wu, PhD,* Jun Chen, MD, PhD,‖§ Stephen N. Thibodeau, PhD,* V. Shane Pankratz, PhD,† and Ping Yang, MD, PhD§*

Introduction: Small cell lung cancer (SCLC) is the most aggressive form of lung malignancy.

Methods: To identify and validate potential DNA methylation markers for risk assessment and disease detection, we examined peripheral blood leukocyte DNA specimens for methylation differences between SCLC cases and controls. We tested 1505 CpG sites using the Illumina Beadchip assay and validated 9 CpG sites using pyrosequencing technology.

Results: In 44 matched SCLC case-control pairs, we identified significant differences at 62 CpG sites (false discovery rate ≤0.05) in 52 independent genes. Of those, we further determined 43 sites in 36 genes with a mean methylation level difference greater than 0.03 between the cases and controls. We then selected and validated 9 CpG sites for methylation differences in an independent set of 138 matched case-control pairs. The 9 validated CpG sites predicted a higher risk for cases than controls in 85.8% of all pairs of cases and controls, and 2 (in genes *CSF3R* and *ERCC1*) jointly contributed most of the discriminating ability.

Conclusions: Our replicated results demonstrated feasibility of applying large-scale methylation arrays for biomarker discovery and subsequent validation in peripheral blood DNA. The CpG sites identified in this study may potentially assist in risk prediction and diagnosis of SCLC.

Key Words: Methylation, Biomarker, Small cell lung cancer, Leukocyte.

(*J Thorac Oncol.* 2010;5: 778–785)

*Department of Laboratory Medicine and Pathology, †Division of Biomedical Statistics and Informatics, ‡Division of Medical Oncology, §Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, College of Medicine, Rochester, Minnesota; and ‖Department of Medical Oncology, The first Affiliated Hospital, China Medical University, 155 North Nanjing Street, Heeping District, Shenyang, PR China.

Disclosure: There are no conflicts of interest by any of the authors. Ping Yang certifies that all authors have agreed to all the content in the manuscript, including the data as presented. In addition, Dr. Yang had full access to all the data in the study and final responsibility for the decision to submit for publication.

Address for correspondence: Ping Yang, MD, PhD, Mayo Clinic, College of Medicine, 200 First Street SW, Rochester, MN 55905. E-mail: yang.ping@mayo.edu

ISSN: 1556-0864/10/0506-0778

Small cell lung cancer (SCLC) constitutes approximately 13% of all newly diagnosed lung cancers.[1] In comparison with the more common non-small cell lung cancer (NSCLC), SCLC has more rapid doubling time, higher growth fraction, earlier development of widespread metastases, and more dramatic initial response to chemotherapy and radiation. Despite high initial responses to therapy, most patients die of recurrent disease. Untreated SCLC has the most aggressive clinical course of any lung tumor, with a median survival of only 2 to 4 months after diagnosis.[1] Cigarette smoking is the strongest risk factor for the development of SCLC. Virtually, all patients with SCLC are current or past smokers, and its risk is related to the duration and intensity of smoking.[2,3]

Epigenetics is defined as the study of heritable changes in gene expression, which occur in the absence of a DNA sequence change and is believed to be important in the etiology of common human diseases[4,5] including cancer. Increasing evidence has demonstrated that epigenetic modifications may result from various types of environmental insults and can lead to cancer development. In a recent review,[6] epigenetics is noted to stand at the epicenter of modern medicine because it unites nuclear reprogramming during development, environmentally induced changes in the body, and the cellular response to external stimuli. Unlike the DNA sequence, epigenetic changes distinguish one tissue type from another, and environmental exposures alter the epigenetic program. The ability of genes to alter their expression is controlled by epigenetic factors such as DNA methylation.[7,8]

Differential methylation status in peripheral blood DNA has been linked to risk of several cancers, although little is known for lung cancer.[9–11] To determine whether global methylation in DNA derived from peripheral blood, an easily accessible tissue, is associated with head and neck squamous cell carcinoma, Hsiung et al[9] assessed the *LRE1* sequence methylation level in a population-based case-control study and found that hypomethylation of the sequence led to a significant 1.6-fold increased risk for the disease. Moore et al[10] reported in a case-control study an association of leukocyte DNA hypomethylation with increased risk of developing bladder cancer, independent of smoking and other assessed risk factors. Currently, Widschwendter et al[11] examined locus-specific methylation and found that particular methylation patterns in peripheral blood DNA may serve as surrogate markers for breast cancer risk. Therefore, methyl-

ation status in peripheral blood DNA specimens may provide a useful biomarker for disease risk assessment and potential early detection and differential diagnosis. Because the methylation status is reversible, further understanding the role of methylation in disease etiology may facilitate targeted therapy and pave the road for future chemoprevention and arresting disease progression. In this study, we used an array-based genomic DNA methylation approach identifying potential quantitative biomarkers for diagnosis or risk assessment of SCLC using peripheral blood DNA in a case-control study.

PATIENTS AND METHODS

Sample Recruitment

The methods of identifying and enrolling patients with SCLC and controls were reported previously.[12–14] In brief, newly diagnosed cases of lung cancer are identified by a daily electronic pathology reporting system. Once identified, patients were consented and enrolled, their medical records were abstracted, and interviews were conducted. Overall participation and blood sample donation rates were 87% and 73%, respectively. For controls, we selected community residents, identified by having had a general medical examination and a leftover blood sample from routine clinical tests, excluding individuals diagnosed with major organ failure (e.g., heart, brain, lung, kidney, or liver) on or before their visit. SCLC cases were identified from among all lung cancer cases, and controls were selected such that the distributions of age, sex, and smoking history were comparable between the cases and controls. Ninety-five percent of the study subjects were white, representing an U.S. Midwestern population in and surrounding Minnesota. The Mayo Clinic Institutional Review Board approved this study.

DNA Modification by Sodium Bisulfite

We extracted DNA from 5 mL of whole blood using an automated platform following QIAGEN kit (Germantown, MD). Because whole blood DNA was predominantly derived from leukocytes and freely circulating DNA in whole blood is negligible, we referred to the whole blood DNA as leukocyte DNA. We modified the genomic DNA specimens using an EZ DNA methylation kit from Zymo Research Corporation (Orange, CA) that combined bisulfite conversion and DNA cleanup. This modification kit is based on the three-step reaction that takes place between cytosine and sodium bisulfite in which cytosine is converted into uracil. We used 1 μg of genomic DNA from peripheral blood DNA for the modification under recommendation from the manufacturer. Treated DNA specimens were stored at -20°C and were assayed within 2 weeks.

Methylation Profiling Analysis

We labeled and hybridized the modified DNA specimens with equal numbers of samples from each group, balanced across the entire Beadchip, to avoid confounding study results with processing variance. We imaged the arrays using a BeadArray Reader scanner, which represented each methylation data point as fluorescent signals from the M (methylated) and U (unmethylated) alleles. The proportion methylated (β value) at each CpG site was calculated using BeadStudio Software (Illumina Inc., San Diego, CA), after subtracting background intensity, computed from negative controls, from each analytical data point.

Pyrosequencing Methylation Assays

Primers were designed using Pyrosequencing Assay Design Software (Biotage AB, Uppsala, Sweden). Sequences of the primers are listed in Table 1. The polymerase chain reaction was carried out on 10 ng of bisulfite-treated DNA using TaqGold DNA polymerase (Applied Biosystems, Foster City, CA) under the following conditions: 10 minutes at 95°C, followed by 50 cycles of 35 seconds at 95°C, 35 seconds at 57.5°C, and 1 minute at 72°C. Pyrosequencing reactions were performed on Biotage PyroMark MD System (Biotage AB) according to the manufacturer's protocols by the sequential addition of single nucleotides in a predefined order. Raw data were analyzed using Pyro Q-CpG 1.0.9 analysis software (Biotage AB). The CpG methylation level (ranging from 0 to 1) was represented by the percentage of methylated C among the sum of methylated and unmethylated C.

Data Analysis

We summarized and compared demographic characteristics between cases and controls using χ^2 tests for nominal variables or rank sum tests for the quantitative variables. We also summarized the percent methylated measurements by their mean and standard deviation within the two study groups and used analysis of covariance approaches to compare the degree of methylation between study groups for each CpG site while adjusting for pack-years of smoking. Because of the non-normality of the methylation values, we used rank-based analyses, which are analogous to rank-sum tests when there are no covariates. After obtaining the p values for each of the CpG sites in the testing set, we used false discovery rate (FDR) approaches and computed a q value for each p value.[15] CpG sites with q values of less than 0.05 were considered to be significant.

In the validation phase, we compared the methylation levels of the 10 selected CpGs between the cases and controls in the validation set using the rank-based procedures outlined earlier. We also used logistic regression approaches to simultaneously assess the association between all nine validation CpGs and case-control status. We further refined this multivariable model by a stepwise model selection with the p value to enter and remain in the model set at 0.05 to determine a CpG set that simultaneously contributes to the discrimination between SCLC cases and controls. As part of these logistic regression analyses, we measured the degree of concordance between model predictions and observed case-control status by extracting estimates of the area under the receiver operating characteristic curve. This quantity, often referred to as the c-statistic, examines all possible case-control pairs and measures the proportion of the time the statistical model predicts higher risk for the case.[16] All analyses were conducted using the SAS software system (Cary, NC).

TABLE 1. Primers for Pyrosequencing Methylation Assay

Genes	Primer Names	Notes	Primer Sequences (From 5′ to 3′)
PECAM1	PECAM1f	PCR-forward, biotin-labeled	TTGAGAAATTAGTTTTGTGAAAAG
	PECAM1r	PCR-reverse	TCAAACCAACCCAAACCCCATTATT
	PECAM1sr	Sequencing-reverse	TTCCAACCATAACTACCATTACCT
S100A2	S100A2f	PCR-forward	GTTAGTTTTATTATTAGTTGGGGGAGGGT
	S100A2r	PCR-reverse, biotin-labeled	ACCCCCATCCAAAATACCC
	S100A2sf	Sequencing-forward	GGAAGTGGGAGGTGT
ERCC1	ERCC1f	PCR-forward	GAGTTAGTGTTGGTGATATAGTAGTGA
	ERCC1r	PCR-reverse, biotin-labeled	CATCCCAAACCTACCCATTCT
	ERCC1sf	Sequencing-forward	TTAAGGTTTAGTAAGGGATATAGATA
SLC22A18	SLC22A18f	PCR-forward	GTGTTTATTTTTAAGATTGGTTGAGGTATT
	SLC22A18r	PCR-reverse, biotin-labeled	TCCCCAACCCCAAAACATT
	SLC22A18sf	Sequencing-forward	TTAGTTAGTTTGGATTTTTTTAT
CSF3R	CSF3Rf	PCR-forward, biotin-labeled	GGGTGTGTTTTAGGTTTTAGGGAATT
	CSF3Rr	PCR-reverse	CCCAAAATTCCTATTTCTCCATCTA
	CSF3Rsr	Sequencing-reverse	CCTAATCTATAAACAATACACAAA
MMP9	MMP9f2	PCR-forward	GTTTGGGGTTTTGTTTGATTTG
	MMP9r2	PCR-reverse, biotin-labeled	CCACCCCTCCTTAACAAACAAATAC
	MMP9seqf2	Sequencing-forward	TGATTTGGTAGTGGAGAT
EMR3	EMR3f2	PCR-forward	ATTTTAGGTTAGTTGATTTATGAAAT
	EMR3r2	PCR-reverse, biotin-labeled	AAATTTACCAACTCAATCATCCCAAAA
	EMR3seqf2	Sequencing-forward	GAAAAGTAAATTGTTTTTTTTTTT
IL-10	IL-10-f1	PCR-forward	TGTAAGTTTAGGGAGGTTTTTTTATTTATT
	IL-10-r1	PCR-reverse, biotin-labeled	AATTCATATTCAACCAATCATTTTTACTT
	IL-10-seqf1	Sequencing-forward	AAGTTATAATTAAGGTTTTT
CAV1	CAV1f3	PCR-forward	AAGGGAAGGTTTAGGATAGGGTAGGATT
	CAV1r3	PCR-reverse, biotin-labeled	TTTTCCCAATACATCATCTCAACA
	CAV1seqfs3	Sequencing-forward	AGGGTAGGATTGTGGAT
TRIP6	TRIP6f2	PCR-forward	GGGTAGGGGTTGGGGAATT
	TRIP6r2	PCR-reverse, biotin-labeled	ATACCCCCCCCCTACTAAACCC
	TRIP6s2	Sequencing-forward	GAAGGGGATTTTGTGA

RESULTS

Characteristics of Study Subjects

By matching design, no difference in age, sex, and smoking status was found between the cases and controls in both the testing and validation sets. Basic descriptive information of the cases and controls are provided in Table 2. For the testing set, five cases were dropped because of DNA quality issues, and the remaining 39 cases and 44 controls were used in the analysis. There was a greater than 3-year difference between the cases and controls in the mean pack-years of cigarette smoking (60.1 versus 56.5). However, median pack-years were similar (51 versus 52), and the test comparing the two groups did not reach statistical significance ($p = 0.525$). To be conservative, the number of pack-years was adjusted in all DNA methylation analyses.

Differentially Methylated CpG Sites

Because the majority of the patients with SCLC received radiation treatment or chemotherapy before blood was drawn, we examined the correlations between the time on treatment (as a proxy for treatment intensity) and the degree of methylation to determine whether the CpG methylation levels might be affected by treatment in the 39 patients with SCLC. Among the 1505 CpG sites, we found that the length of time on treatment was significantly correlated with the methylation levels of 173 CpGs ($p < 0.05$). Although some of these associations may be false-positives, to be conservative, we excluded all 173 CpGs from the analyses. Among the remaining 1332 CpG sites, 922 were located within CpG islands and 410 were in non-CpG islands. We observed significant differences between SCLC cases and controls at 62 sites in 52 independent genes (FDR $\leq$0.05). Interestingly, only 25 of the 62 sites were in CpG islands, which was significantly lower than the expected 42.9 sites ($62 \times 922/1332$) ($p < 0.001$, χ^2 test). The odds of a significant CpG not being in a CpG island was greater than three times higher than the odds of being in a CpG island (OR = 3.56, 95% confidence interval [CI]: 2.11–6.00). Furthermore, only 6 of the 62 sites showed an increased level of methylation in patients with SCLC, including 2 in the *ITK* gene, 2 in the *RUNX3* gene, 1 in each of the *CTLA4* and *PLG* genes. Because some methylation differences were small and difficult to reliably detect, we further excluded the CpG sites with an absolute mean β difference of less than 0.03, resulting in 43 significant CpG sites of primary interest in 36 independent genes (Table 3).

TABLE 2. Characteristics of Patients with SCLC and Healthy Controls

	Testing Set				Validation Set			
	Cases (n = 39)	Controls (n = 44)	Total (N = 83)	p	Cases (n = 138)	Controls (n = 138)	Total (N = 276)	p
Age				0.5392				0.6736
Mean (SD)	64.8 (6.1)	65.6 (5.7)	65.2 (5.8)		64.4 (9.54)	64.8 (9.46)	64.6 (9.48)	
Median	65.0	65.5	65.0		66.0	66.0	66.0	
Q1, Q3	61.0, 69.0	61.5, 69.0	61.0, 69.0		59.0, 71.0	59.0, 71.0	59.0, 71.0	
Range	54.0–78.0	57.0–78.0	54.0–78.0		37.0–85.0	33.0–82.0	33.0–85.0	
Gender				0.6470				1
Female	17 (43.6%)	17 (38.6%)	34 (41.0%)		61 (44.2%)	61 (44.2%)	122 (44.2%)	
Male	22 (56.4%)	27 (61.4%)	49 (59.0%)		77 (55.8%)	77 (55.8%)	154 (55.8%)	
Smoking status				0.9072				1
Never	0	0	0		8 (5.8%)	8 (5.8%)	16 (5.8%)	
Former	20 (51.3%)	22 (50.0%)	42 (50.6%)		63 (45.7%)	63 (45.7%)	126 (45.7%)	
Current	19 (48.7%)	22 (50.0%)	41 (49.4%)		67 (48.6%)	67 (48.6%)	134 (48.6%)	
Pack-years				0.5249				0.9218
Mean (SD)	60.1 (25.1)	56.5 (25.8)	58.2 (25.4)		56.8 (29.4)	56.4 (29.2)	56.6 (29.3)	
Median	51.0	52.0	52.0		52.0	52.0	52.0	
Q1, Q3	42.0, 74.0	39.0, 68.3	41.0, 72.0		37.0, 75.0	36.0, 77.0	36.0, 76.5	
Range	22.0–126.0	17.0–141.0	17.0–141.0		3.0–146.0	3.0–147.0	3.0–147.0	

Validation of Selected CpG Sites by Pyrosequencing Methylation Assay

Based on three major parameters (FDR q values, number of significant CpGs/gene, and mean difference between groups), we selected 10 CpG sites including 9 significant CpGs (FDR <0.05) for validation and 1 nonsignificant CpG (FDR >0.05). These CpG sites were located in 10 different genes (*IL10, PECAM1, S100A2, MMP9, ERCC1, EMR3, SLC22A18, TRIP6, CSF3R,* and *CAV1*), with *CAV1* serving as a negative control. We designed a new assay for each of the 10 CpG sites using pyrosequencing technology.[17,18] Figure 1 shows methylation levels of a CpG site, 85bp upstream to the transcription start site in the gene, *IL10*, in three different samples.

We then tested the 10 CpG sites for methylation differences, again in peripheral blood DNA specimens from a validation set between 138 SCLC cases and 138 matched controls (Table 2, right panel). The nine testing-set-positive CpG sites again demonstrated significant differences (all p values ≤ 0.0003, Table 4), whereas the negative control CpG site only differed between the validation set of the cases and controls in an absolute percent methylated by less than 1%. This small difference did not reach statistical significance.

CpG Methylation Patterns and Risk Prediction of SCLC Using Logistic Regression Models

Based on the nine validated CpG sites accounting for age, sex, and smoking history, our model had an area under the receiver operating characteristic curve of 0.858 (Figure 2), suggesting the model correctly classified SCLC cases as being at a higher risk than controls for 85.8% of case-control pairs. Further stepwise selection identified two of the nine sites, one in *CSF3R* and the other in *ERCC1*, contributing independent information to discriminate cases from controls. Specifically, for each 5% decrease in the methylation level of *ERCC1*, there was an approximately 4-fold (OR = 3.9, 95% CI: 2.0–6.1, $p < 0.001$) increase in the odds ratio of SCLC; for each 5% methylation decrease of *CSF3R,* there was a 1.5-fold higher OR of SCLC (OR = 1.5, 95% CI: 1.1–2.0, $p = 0.008$).

DISCUSSION

In applying newly developed methylation BeadChip technology, we performed methylation profiling analysis of 1505 tested CpG sites in 807 genes and identified 62 CpGs whose methylation status were significantly associated with SCLC. Forty-three of these 62 had an absolute mean difference greater than 0.03 between the cases and controls, and 9 of these were further confirmed using pyrosequencing methylation assays in additional cases and controls. These results suggest that methylation status of peripheral blood DNA, a stable and easily accessible material, might be reliably used for risk assessment and diagnosis of SCLC.

Among the nine validated genes, at least five were reported to be associated with lung cancer. For example, the *IL10* expression by tumor-associated macrophages showed a significant role in the progression and prognosis of NSCLC.[19] Patients whose tumors had a positive S100A2 expression had a significantly lower overall survival and disease-specific survival rate.[20] Another study showed S100A2 being upregulated in an early stage of patients with NSCLC, indicating its important role for molecular diagnosis of NSCLC at an early stage; therefore, prognostically more favorable.[21] High ERCC1 expression was associated with short survival.[22,23] Interestingly, results of these studies came from cancer tissues and were consistent with our findings from peripheral blood DNA, indirectly supporting that hypomethylation of

TABLE 3. Differential Methylations Between 39 SCLC Cases and 44 Healthy Controls in Testing Set

Symbol	Illumina CpG ID[a]	CpG Island	Adjusted *p*	*q*	Controls		Cases		Case/Control Difference
					Mean β	SD	Mean β	SD	
SLC22A18	**SLC22A18_P216_R**	No	0.00002	0.00877	0.464	0.091	0.354	0.131	−0.11
PADI4	PADI4_E24_F	No	0.00002	0.00877	0.257	0.067	0.190	0.096	−0.067
MMP9	**MMP9_P189_F**	No	0.00005	0.00877	0.377	0.076	0.298	0.100	−0.079
LTB4R	LTB4R_P163_F	No	0.00005	0.00877	0.303	0.063	0.242	0.068	−0.061
S100A2	**S100A2_E36_R**	No	0.00006	0.00877	0.353	0.070	0.282	0.073	−0.071
RUNX3	RUNX3_P247_F	Yes	0.00006	0.00877	0.703	0.102	0.790	0.153	**0.087**
MPO	MPO_E302_R	No	0.00007	0.00877	0.700	0.065	0.618	0.089	−0.082
IL10	**IL10_P85_F**	No	0.00007	0.00877	0.229	0.051	0.178	0.076	−0.051
RUNX3	RUNX3_E27_R	No	0.00008	0.00885	0.862	0.049	0.900	0.092	**0.038**
PECAM1	**PECAM1_E32_R**	Yes	0.00008	0.00885	0.257	0.065	0.189	0.087	−0.068
EMR3	**EMR3_E61_F**	No	0.00014	0.01292	0.22	0.049	0.166	0.085	−0.054
SPI1	SPI1_E205_F	Yes	0.00017	0.01292	0.315	0.057	0.250	0.100	−0.065
TNFRSF1A	TNFRSF1A_P678_F	No	0.00019	0.01292	0.754	0.070	0.662	0.125	−0.092
LMO2	LMO2_E148_F	No	0.00019	0.01292	0.442	0.103	0.327	0.151	−0.115
IL10	IL10_P348_F	No	0.0002	0.01292	0.64	0.086	0.509	0.180	−0.131
ERCC1	**ERCC1_P440_R**	Yes	0.0002	0.01292	0.141	0.046	0.103	0.047	−0.038
CSF3R	**CSF3R_P472_F**	No	0.00041	0.02392	0.371	0.097	0.276	0.134	−0.095
JAK3	JAK3_P1075_R	No	0.00044	0.02392	0.683	0.077	0.617	0.087	−0.066
LCN2	LCN2_P141_R	No	0.00048	0.02392	0.789	0.078	0.708	0.131	−0.081
CD82	CD82_P557_R	Yes	0.00048	0.02392	0.277	0.097	0.192	0.109	−0.085
PI3	PI3_P274_R	No	0.00052	0.02457	0.835	0.053	0.763	0.110	−0.072
TRIP6	**TRIP6_P1090_F**	Yes	0.00053	0.02457	0.359	0.107	0.259	0.139	−0.100
TIE1	TIE1_E66_R	No	0.00055	0.02457	0.224	0.067	0.164	0.085	−0.06
TRIP6	TRIP6_P1274_R	Yes	0.00061	0.02543	0.617	0.101	0.488	0.178	−0.129
CD9	CD9_P585_R	Yes	0.00063	0.02548	0.385	0.061	0.314	0.103	−0.071
SEPT9	SEPT9_P374_F	Yes	0.00065	0.02555	0.252	0.097	0.180	0.102	−0.072
MPL	MPL_P62_F	No	0.00091	0.0319	0.492	0.098	0.389	0.151	−0.103
CASP10	CASP10_P334_F	No	0.00094	0.0319	0.243	0.059	0.191	0.078	−0.052
AIM2	AIM2_P624_F	No	0.00094	0.0319	0.467	0.137	0.353	0.178	−0.114
SEPT9	SEPT9_P58_R	Yes	0.001	0.03234	0.929	0.046	0.888	0.061	−0.041
CSF1R	CSF1R_E26_F	No	0.00107	0.03239	0.749	0.075	0.662	0.135	−0.087
CSF3R	CSF3R_P8_F	No	0.0013	0.03635	0.225	0.072	0.168	0.096	−0.057
MMP14	MMP14_P13_F	Yes	0.00167	0.04137	0.500	0.101	0.400	0.157	−0.100
BTK	BTK_P105_F	No	0.00167	0.04137	0.132	0.047	0.100	0.050	−0.032
GRB7	GRB7_E71_R	No	0.00178	0.04319	0.374	0.092	0.296	0.134	−0.078
STAT5A	STAT5A_P704_R	No	0.00189	0.04428	0.183	0.06	0.139	0.059	−0.044
NOTCH4	NOTCH4_E4_F	No	0.00189	0.04428	0.170	0.069	0.123	0.072	−0.047
HOXB2	HOXB2_P99_F	Yes	0.00214	0.04667	0.552	0.09	0.483	0.103	−0.069
MFAP4	MFAP4_P197_F	No	0.0022	0.04721	0.244	0.061	0.196	0.080	−0.048
SLC5A5	SLC5A5_E60_F	Yes	0.00227	0.04721	0.527	0.089	0.459	0.107	−0.068
CD34	CD34_P339_R	No	0.00227	0.04721	0.268	0.049	0.236	0.056	−0.032
MFAP4	MFAP4_P10_R	No	0.00249	0.04933	0.172	0.061	0.133	0.059	−0.039
EMR3	EMR3_P39_R	No	0.00264	0.04933	0.287	0.064	0.232	0.072	−0.055
CAV1	**CAV1_P169_F**[b]	Yes	0.35439	0.63028	0.191	0.056	0.176	0.054	−0.015

[a] The CpG IDs in bold are selected to run pyrosequencing for validation. The CpG site in the gene *CAV1* was selected as negative control. SD, standard deviation.

these genes may increase gene expression. To date, it is not known whether the abnormal methylation in peripheral blood DNA from patients with SCLC is inherited or acquired. Further investigation is clearly needed.

Imprinting genes in peripheral blood leukocytes may also serve as an indicator of cancer risk.[24] In this study, we observed an imprinting gene, *SLC22A18,* whose methylation levels in the tested CpG demonstrated a strong association with SCLC in both the testing and validation sets (adjusted $p < 0.0001$, Tables 3 and 4). The gene is located in the imprinted gene domain of 11p15.5, an important tumor-suppressor gene region; alterations in this region have been associated with the Beckwith-Wiedemann syndrome and multiple malignancies including lung cancer.[25–27] The finding

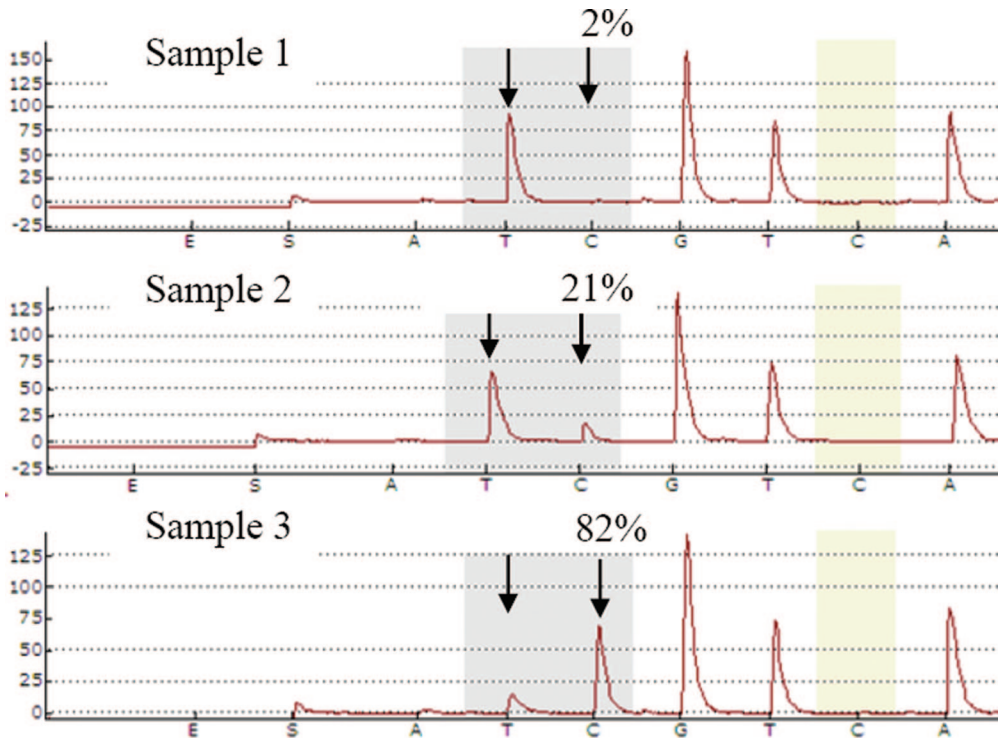

FIGURE 1. Methylation analysis of IL10_P85_F CpG by pyrosequencing technology. Methylation analysis of IL10_P85_F CpG in three representative samples by pyrosequencing technology. The first arrow at the T position indicates signal peak of unmethylated C. The second arrow at the C position represents signal peak of methylated C. The CpG methylation level is the percentage of methylated C among the sum of methylated C and unmethylated C. The methylation levels for the samples 1, 2, and 3 are 2%, 21%, and 82%, respectively.

TABLE 4. Differential Methylations Between 138 SCLC Cases and 138 Matched Controls: Validation Study

Gene Symbol	Illumina CpG IDs	CpG Island	Adjusted *p*	Control		Case		Case/Control Difference
				Mean Methylation Level	SD	Mean Methylation Level	SD	
IL10	IL10_P85_F	No	<0.0001	0.116	0.032	0.077	0.035	−0.039
PECAM1	PECAM1_E32_R	Yes	<0.0001	0.343	0.089	0.242	0.104	−0.101
S100A2	S100A2_E36_R	No	<0.0001	0.288	0.076	0.211	0.063	−0.077
MMP9	MMP9_P189_F	No	<0.0001	0.058	0.020	0.037	0.021	−0.021
ERCC1	ERCC1_P440_R	Yes	<0.0001	0.135	0.044	0.085	0.037	−0.050
EMR3	EMR3_E61_F	No	<0.0001	0.161	0.043	0.115	0.050	−0.046
SLC22A18	SLC22A18_P216_R	No	<0.0001	0.227	0.059	0.155	0.074	−0.072
TRIP6	TRIP6_P1090_F	Yes	0.0003	0.44	0.255	0.319	0.259	−0.121
CSF3R	CSF3R_P472_F	No	<0.0001	0.277	0.066	0.18	0.084	−0.097
CAV1	CAV1_P169_F	Yes	0.3577	0.1	0.063	0.109	0.07	0.009

SD, standard deviation.

that the *SLC22A18* methylation level is associated with SCLC further demonstrates the importance of the imprinting region in cancer etiology.

Global hypomethylation has been reported in cancer tissues (somatic) and constitutive tissues (germline).[9,10,28–30] Lack of DNA methylation at specific CpG sites has also been reported recently to be associated with breast cancer risk.[11] Of the 62 CpG sites that were identified in this study, 56 are hypomethylated and only 6 are hypermethylated. The hypomethylation may be caused by different mechanisms. First, tumor formation requires activation of oncogenes and tumor response genes, which can be achieved through promoter hypomethylation of related genes. Second, this hypomethylation may be part of global hypomethylation commonly seen in cancer patients and may be not associated with activation of any specific genes. The fact that methylation changes preferentially occur in non-CpG islands indirectly support this possibility. If global hypomethylation causes genomic instability[29]

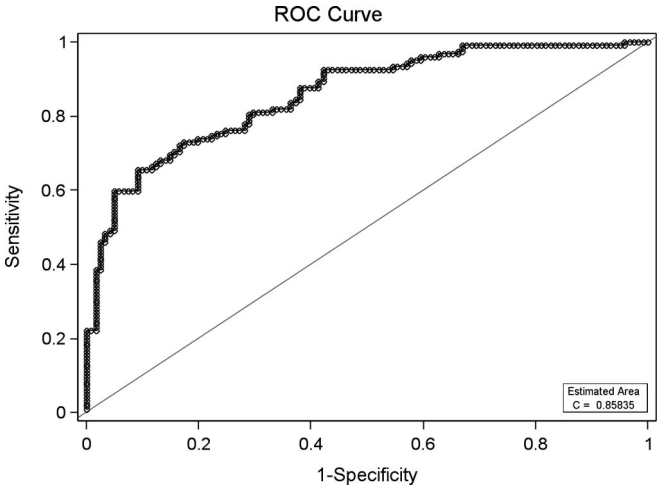

FIGURE 2. ROC curve in the validation set of 138 cases and 138 controls using the nine CpGs selected from the set. The curve illustrates the capacity of the methylation levels of the CpGs to discriminate between SCLC cases and controls. Of note, the area under the ROC curve (the c-statistic), represents the proportion of SCLC case-control pairs in which the case is predicted by the model based on the nine CpGs to have greater odds of being a SCLC case. ROC, receiver operating characteristic.

and hence increases risk to cancer, the current finding from peripheral blood DNA may represent, at least in part, constitutive hypomethylation of patients with SCLC and explain why these individuals are susceptible to developing cancer.

Among the six hypermethylated CpG sites, two are from the gene *RUNX3*. The gene, a tumor suppressor, is inactivated in a variety of cancers including SCLC.[31–34] The promoter hypermethylation of the gene is believed to be a major mechanism for the gene inactivation in tumor tissues. One of the two CpG sites (RUNX3_P247_F) is located within a CpG island (promoter region) and shows an 8.72% difference between SCLC cases and controls. Although not tested, we predict a lower activity of *RUNX3* in peripheral blood leukocytes of patients with SCLC. If reduced activity of the gene is also seen in lung tissue, it may be responsible for the increased risk of SCLC.

However, this study has some limitations. First, we were unable to determine what mechanisms caused the differential methylation between the cases and controls. The possible mechanisms include differences in genetics, treatment, nutritional status, and subpopulation of leukocytes. A comprehensive analysis that addresses these potential mechanisms will be necessary for future study. Second, we tested only one to two CpG sites per gene that are predefined by the manufacturer for inclusion on the methylation array that we used (Illumina Inc.). The tested CpGs are not necessarily most representative for a particular gene. More detailed analysis is needed to fully cover all CpGs in related genes. Nevertheless, our results demonstrate methylation differences between patients with SCLC and controls are present and can be reliably detected in peripheral blood leukocyte DNA. SCLC is traditionally not treated with surgical resection; thus, it is difficult to obtain adequate tissue samples for molecular analysis. DNA from peripheral blood leukocytes has been widely used in genetic analysis. The successful use of the easily accessible specimen in this study will significantly expand the research application from genetics (such as genome-wide association studies) to epigenetics (such as epigenome-wide association studies). We believe that our methylation panel is potentially useful as a second-tier disease prediction and a noninvasive detection tool among high-risk individuals, particularly smokers with equivocal findings from computed tomography screening.

ACKNOWLEDGMENTS

This work was partly supported by the National Institutes of Health (NIH) research grants: R01 CA 80127, R01 CA 84354, and R03 CA 77118.

We thank Susan Ernst, MA, for her technical assistance and Dr. Zhifu Sun for his helpful suggestions.

REFERENCES

1. Stupp R, Monnerate C, Turrisi AT. Small cell lung cancer: state of the art and future perspectives. *Lung Cancer* 2004;45:105–117.
2. Brownson RC, Chang JC, Davis JR. Gender and histologic type variations in smoking-related risk of lung cancer. *Epidemiology* 1992;3:61–64.
3. de Andrade M, Ebbert JO, Wampfler JA, et al. Environmental tobacco smoke exposure in women with lung cancer. *Lung Cancer.* 2004;43: 127–134.
4. Jiang YH, Bressler J, Beaudet AL. Epigenetics and human disease. *Ann Rev Genomics Human Genet* 2004;5:479–510.
5. Bjornsson HT, Fallin MD, Feinberg AP. An integrated epigenetic and genetic approach to common human disease. *Trends Genet* 2004;20: 350–358.
6. Feinberg AP. Epigenetics at the epicenter of modern medicine. *J Natl Cancer Inst* 2008;299:1345–1350.
7. Dolinoy DC, Jirtle RL. Environmental epigenomics in human health and disease. *Environ Mol Mutagen* 2008;49:4–8.
8. Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 2007;447:425–432.
9. Hsiung DT, Marsit CJ, Houseman EA, et al. Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev* 2007;16:108–114.
10. Moore LE, Pfeiffer RM, Poscablo C, et al. Genomic DNA hypomethylation as a biomarker for bladder cancer susceptibility in the Spanish Bladder Cancer Study: a case-control study. *Lancet Oncol* 2008;9:359–366.
11. Widschwendter M, Apostolidou S, Raum E, et al. Epigenotyping in peripheral blood cell DNA and breast cancer risk: a proof of principle study. *PLoS One.* 2008;3:e2656.
12. Yang P, Sun Z, Krowka MJ, et al. Alpha-1 antitrypsin deficiency carriers, tobacco smoke, chronic obstructive pulmonary disease, and lung cancer risk. *Arch Int Med* 2008;168:1097–1103.
13. Yang P, Allen MS, Aubry MC, et al. Clinical features of 5,628 primary lung cancer patients: experience at Mayo Clinic from 1997–2003. *Chest* 2005;128:452–462.
14. Yang P, Wentzlaff KA, Katzmann JA, et al. Alpha$_1$-antitrypsin deficiency allele carriers in lung cancer patients. *Cancer Epidemiol Biomarkers Prev* 1999;8:461–465.
15. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003;100:9440–9445.
16. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39: 561–577.
17. Tost J, Gut IG. Analysis of gene-specific DNA methylation patterns by pyrosequencing technology. *Methods Mol Biol* 2007;373:89–102.
18. Tost J, Gut IG. DNA methylation analysis by pyrosequencing. *Nat Protoc.* 2007;2:2265–2275.

19. Zeni E, Mazzetti L, Miotto D, et al. Macrophage expression of interleukin-10 is a prognostic factor in nonsmall cell lung cancer. *Eur Respir J* 2007;4:627–632.
20. Wang H, Zhang Z, Li R, et al. Overexpression of S100A2 protein as a prognostic marker for patients with stage I non small cell lung cancer. *Int J Cancer* 2008;116:285–290.
21. Bartling B, Rehbein G, Schmitt WD, et al. S100A2-S100P expression profile and diagnosis of non-small cell lung carcinoma: impairment by advanced tumour stages and neoadjuvant chemotherapy. *Eur J Cancer* 2007;43:1935–1943.
22. Ota S, Ishii G, Goto K, et al. Immunohistochemical expression of BCRP and ERCC1 in biopsy specimen predicts survival in advanced non-small-cell lung cancer treated with cisplatin-based chemotherapy. *Lung Cancer* 2009;64:98–104.
23. Okuda K, Sasaki H, Dumontet C, et al. Expression of excision repair cross-complementation group 1 and class III beta-tubulin predict survival after chemotherapy for completely resected non-small cell lung cancer. *Lung Cancer* 2008;62:105–112.
24. Cui H, Cruz-Correa M, Giardiello M, et al. Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science* 2003;299:1753–1755.
25. Kohda M, Hoshiya H, Katoh M, et al. Frequent loss of imprinting of IGF2 and MEST in lung adenocarcinoma. *Mol Carcinog* 2001;31:184–191.
26. Kondo H, Suzuki H, Ueda R, et al. Frequent loss of imprinting of the h19 gene is often associated with its overexpression in human lung cancers. *Oncogene* 1995;10:1193–1198.
27. Scott RH, Douglass J, Baskcomb, et al. Constitutional 11p15 abnormalities, including heritable imprinting center mutations, cause nonsyndromic Wilms tumor. *Nat Genet* 2008;40:1329–1334.
28. Chalitchagorn K, Shuangshoti S, Hourpai N, et al. Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis. *Oncogene* 2004;23:8841–8846.
29. Daskalos A, Nikolaidis G, Xinarianos G, et al. Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *Int J Cancer* 2009;124:81–87.
30. Ogino S, Nosho K, Kirkner GJ, et al. A cohort study of tumoral LINE-1 hypomethylation and prognosis in colon cancer. *J Natl Cancer Inst* 2008;100:1734–1738.
31. Suzuki M, Shigematsu H, Shames DS, et al. DNA methylation-associated inactivation of TGF-beta-related genes DRM/Gremlin, RUNX3, and HPP1 in human cancers. *Br J Cancer* 2005;93:1029–1037.
32. Sato K, Tomizawa Y, Iijima H, et al. Epigenetic inactivation of the RUNX3 gene in lung cancer. *Oncol Rep* 2006;15:129–135.
33. Li QL, Kim HR, Kim WJ, et al. Transcriptional silencing of the RUNX3 gene by CpG hypermethylation is associated with lung cancer. *Biochem Biophys Res Commun* 2004;314:223–228.
34. Yanada M, Yaoi T, Shimada J, et al. Frequent hemizygous deletion at 1p36 and hypermethylation downregulate RUNX3 expression in human lung cancer cell lines. *Oncol Rep* 2005;14:817–822.

